ZOOM AI SIGN LANGUAGE INTERPRETATION

DESIGN PROJECT-II

Submitted by

R E Dharshan – 21113049

G Madhulika Reddy – 21113073

**Under the guidance of**

**PRAISY EVANGELIN A**

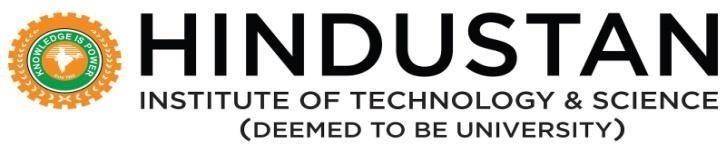
**Professor**

***in partial fulfillment for the award of the degree of***

**BACHELOR OF TECHNOLOGY**

**in**

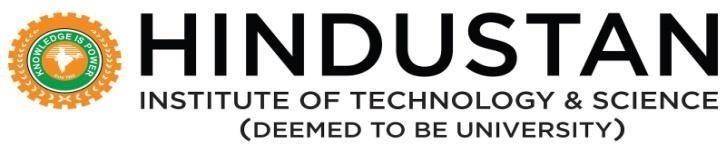
**COMPUTER SCIENCE AND ENGINEERING**



**HINDUSTAN INSTITUTE OF TECHNOLOGY AND SCIENCE**

**CHENNAI - 603 103**

MAY 2024



**BONAFIDE CERTIFICATE**

Certified that this project report **ZOOM AI SIGN LANGUAGE INTERPRETATION** is the bonafide work of **R E Dharshan 21113049, G Madhulika Reddy 21113073** who carried out the project work under my supervision during the academic year 20**23**-20**24**.

PRAISY EVANGELIN A,

**SUPERVISOR**

PROFESSOR

|  |  |  |
| --- | --- | --- |
|  |  |  |
| **INTERNAL EXAMINER** |  | **EXTERNAL EXAMINER** |
| Name:  Designation: |  | Name:  Designation: |

Department of CSE.

Project Viva - voce conducted on :

**Contents**

[Table of Figures iv](#_Toc19735)

[Acknowledgement v](#_Toc19736)

[Abstract vi](#_Toc19737)

[CHAPTER 1 7](#_Toc19738)

[1.1. Introduction 7](#_Toc19739)

[1.2. Motivation of the work 7](#_Toc19740)

[CHAPTER 2 9](#_Toc19741)

[2.1. Introduction 9](#_Toc19742)

[2.2. Literature Survey 9](#_Toc19743)

[2.3. Summary 10](#_Toc19744)

[CHAPTER 3 11](#_Toc19745)

[3.1. Model Description 11](#_Toc19746)

[3.1.1 Introduction 11](#_Toc19747)

[3.1.2 Architecture Diagram 11](#_Toc19748)

[3.2. Data Sets 12](#_Toc19749)

[3.3. Tools 12](#_Toc19750)

[3.4. Summary 13](#_Toc19751)

[CHAPTER 4 14](#_Toc19752)

[4.1. Experimental setup analysis 14](#_Toc19753)

[4.1.1 GPT-2 for Text Summarization 14](#_Toc19754)

[4.2. Results 15](#_Toc19755)

[4.3. Summary 16](#_Toc19756)

[CHAPTER 5 17](#_Toc19757)

[5.1. Conclusion 17](#_Toc19758)

[5.2. Future Enhancements 18](#_Toc19759)

[Appendix A 19](#_Toc19760)

[Appendix B 20](#_Toc19761)

[Appendix C 22](#_Toc19762)

[References 23](#_Toc19763)

# Table of Figures

Figure 1: Architecture Diagram 11

Figure 2: Training Data Set 12

Figure 3: Test data which has to be summarized 14

Figure 4: Screenshot 15

Figure 5: Train and Test Score 15

Figure 6: ROUGE Metrics 16

# ACKNOWLEDGEMENT

First and foremost, we would like to thank **ALMIGHTY** who has provided us the strength to do justice to our work and contribute our best to it.

We wish to express our deep sense of gratitude from the bottom of our heart to our Guide **Praisy Evangelin A, Professor**, **Computer Science and Engineering,** for her motivating discussions, overwhelming suggestions, ingenious encouragement, invaluable supervision, and exemplary guidance throughout the project work.

We would like to extend our heartfelt gratitude to **Dr. J.Thangakumar,**

**Ph.D., Professor & Head, Department of Computer Science and Engineering** for his valuable suggestions and support in successfully completing the project. We are really indebted to our **Dean Computing Sciences Dr.R.Kanthavel,** for his support.

We thank the management of **HINDUSTAN INSTITUTE OF TECHNOLOGY AND SCIENCE** for providing us the necessary facilities and support required for the successful completion of the project.

As a final word, we would like to thank each and every individual who have been a source of support and encouragement and helped us to achieve our goal and complete our project work successfully.

# 

# ABSTRACT

This report investigates the efficacy of employing GPT-2, a state-of-theart language model, for text summarization tasks. Leveraging the advancements in natural language processing, GPT-2 demonstrates remarkable capabilities in understanding and generating human-like text. The study employs a dataset of diverse textual sources, to evaluate the performance of GPT-2 in summarizing lengthy documents into concise and coherent summaries. Through extensive experimentation and evaluation metrics, including ROUGE scores and human evaluation, the report examines the quality and effectiveness of the generated summaries. Furthermore, the report discusses the strengths, limitations, and potential applications of utilizing GPT-2 for text summarization in various domains. The findings contribute to a deeper understanding of the capabilities and implications of employing advanced language models for automating the summarization process, thereby paving the way for enhanced information retrieval and comprehension in the digital era. This approach to summarization offers several advantages over traditional methods, including the ability to capture semantic nuances, generate human-like summaries, and handle a wide range of input formats and topics.

# CHAPTER 1

## 1.1. Introduction

Video conferencing platforms have become essential tools for social interaction, remote communication, and teamwork in the current digital era. However, the lack of strong accessibility features in current platforms and the scarcity of sign language interpreters pose particular challenges for people who are deaf or hard of hearing when participating in video calls. We suggest a ground-breaking project that aims to incorporate artificial intelligence (AI) technologies for real-time sign language interpretation directly within the Zoom video conferencing platform in order to address this important issue and foster inclusive communication.

This project's main objective is to create a seamless, user-friendly solution that lets sign language users fully participate in Zoom meetings, webinars, and online events. Our goal is to give Zoom users the capacity to enable real-time sign language interpretation during video calls by utilizing computer vision and machine learning techniques. This will improve accessibility and promote inclusivity in online communication.

This project was inspired by the realization that people with hearing loss encounter communication obstacles when interacting in digital settings. The availability, scalability, and cost-effectiveness of traditional solutions—like employing human interpreters or depending on outside assistive technologies—are frequently constrained. Our goal is to overcome these constraints by utilizing AI, and to offer a scalable and effective substitute that smoothly fits into current video conferencing processes.

Creating reliable deep learning models that can precisely identify and interpret sign language gestures recorded in real-time video streams is a crucial part of our suggested solution. To guarantee that a wide range of sign vocabulary and gestures across various languages and dialects are covered, these models will be trained on extensive sign language datasets. In addition, we will create and deploy a unique Zoom plugin that connects to our AI interpreter backend, allowing Zoom users to switch on sign language interpretation during meetings.

We will improve accuracy, performance, and user experience by iterating and fine-tuning our integrated solution based on user feedback and usability testing. Our ultimate goal is to enable anyone with a variety of communication needs to participate meaningfully in online discussions and teamwork by democratizing access to sign language interpretation within the Zoom ecosystem.

To sum up, by utilizing AI-driven sign language interpretation, this project marks a significant advancement in video conference accessibility standards. Our goal is to create a more inclusive digital society and enable universal communication access by directly integrating inclusive technologies into popular platforms like Zoom.

## 1.2. Motivation of the work

This project was motivated by the urgent need to address the communication barriers that hard of hearing or deaf people encounter in virtual environments, especially during video conferences. Even though Zoom and other similar platforms are widely used for remote communication, access to sign language interpretation is still scarce and frequently depends on human interpreters who might not always be available.

The project's goal is to improve inclusivity and accessibility in video conference settings by utilizing artificial intelligence's (AI) transformative potential. We aim to enable people who use sign language as their primary form of communication to fully and independently participate in virtual meetings, webinars, and collaborative sessions by directly integrating AI-driven sign language interpretation into Zoom.

Conventional approaches, like using external assistive technologies or employing interpreters on-site, can be expensive, inconvenient, and unable to grow to accommodate the varied needs of a worldwide audience. By delivering real-time, on-demand sign language interpretation that is seamlessly integrated into the Zoom user experience, leveraging AI technologies presents a promising alternative.

The project is also driven by the way inclusive technologies have the potential to advance universal access to communication. We hope to reduce barriers to communication, encourage equal participation, and advance an inclusive digital society where people with a range of communication needs can confidently and meaningfully engage by democratizing access to sign language interpretation within a popular platform such as Zoom.

In the end, this project aims to show how AI-driven accessibility features can be incorporated into mainstream technologies and show their benefits. It also advocates for the adoption of inclusive design principles, which guarantee that digital platforms are usable by all users, irrespective of their communication preferences or abilities. We hope to advance accessibility standards in virtual communication and pave the way for a more inclusive future for all through research, development, and user-centered design.

# CHAPTER 2

## 2.1. Introduction

Through the analysis of hand movements, gestures, and facial expressions, our system can dynamically interpret sign language in real-time, providing instant translation for all participants in a virtual meeting. Furthermore, our solution is designed to be accessible and user-friendly, requiring minimal setup and configuration. By leveraging standard devices such as webcams and smartphones, our system eliminates the need for specialized hardware, making it more accessible and cost-effective for both individuals and organizations.

.

## 2.2. Literature Survey

A lot of research has been done on sign language interpretation—both technologically and manually driven—to improve accessibility for people who are hard of hearing or deaf. Although conventional approaches depend on human translators, new developments in AI-driven methodologies, especially in deep learning techniques, present encouraging substitutes. In sign language recognition challenges, deep learning models like convolutional neural networks (CNNs) and recurrent neural networks (RNNs) have demonstrated efficacy, opening the door for automatic interpretation from sign language input.

##### Table 1: Literature survey

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Title** | **Author** | **Publishing Year** | **Algorithms Used** | **Findings** |
| Review paper on sign language recognition for the deaf and dumb | R. Rumana ,& R . Prema | 2023 | Analyze video using hand detection, feature extraction and classification with traditional machine learning. | Accuracy:80%    Often lack sufficient data for less common signs or regional variations within sign languages |
| American Sign Language Words Recognition Using Spatio-Temporal Prosodic and Angle Features: A Sequential Learning Approach  Sign Language Recognition via Late Fusion of Computer Vision and Leap Motion | B.A. Sunusi & C. Kosin  Jordan J. Bird 1, Anikó Ekárt and Diego R. Faria | 2022  2022 | FFV-Bi-LSTM to train 3D hand skeletal information of motion and orientation angle features learned from the leap motion controller (LMC).  Computer Vision Feature Extraction and Leap Motion Data Acquisition  Late Fusion Integration and Gesture Classification | FFV is trial and error strategy while choosing stable GMM components.  Limited Leap Motion Range  Environmental factors affect accuracy |
| Accuracy Enhancement of Hand Gesture  Recognition Using CNN | Gyu Tae Park ,  & V. K Chandrasekar | 2023 | Utilizing CNN architecture with diverse datasets and optimized training techniques enhances hand gesture recognition accuracy. | Variability in hand gestures Limited dataset diversity |
| A Review of the Hand Gesture Recognition  System | Noraini Mohamed ,& Nazean Johmar | 2021 | Histogram of Oriented Gradient (HOG), Convolutional Neural Network (CNN), and Principal Component Analysis (PCA). | It needs to work with sign variations. It is difficult to be addressed in practice, especially for vision-based systems because of the related constraints. |
| Intelligent Sign Language Recognition Using Image Processing | Sawant Pramada, & Nerkar Samiksh,& S. Vaidya | 2021 | Image Processing and Template matching for better output generation. | These systems may not be able to recognize all signs, especially rare or regional signs. |
| Deep Learning-Based Standard Sign  Language Discrimination | Menglon Zhang ,& Min Zhao | Institute of Electrical and Electronics Engineers Journal(IEEE), 2023 | the FGSFP+TFFR was fine-tuned with pretrained key frame detection model.  the DCSR3D+GRU model was designed to realize comprehensive correctness discrimination of the sign language category and standardization. | Small datasets make it difficult for models to learn complex patterns.   Imbalances in the dataset can lead to biased recognition results. |
| Real Time Sign Language Interpreter | G.N. Geethu ,& C.S. Arun | International Journal on Electrical, Instrumentation and Communication Engineering,  2022 | Hand sign recognition system was implemented using ARM CORTEX development board. | Low-resolution images or inconsistent signing can impact accuracy. |
| Sign Language Recognition Using Deep  Learning and Computer Vision | Dr. Sabeenian R.S | 2021 | a CNN based approach for the recognition and classification of the sign language using computer vision. | it doesn’t support background subtraction when the frames are dropped from a video. |
| Dynamic Korean Sign Language Recognition | Jung Pil Shin ,& A.S. Musa  ,& K. Suzuki | 2023 | GCN and an attention-driven neural framework, resulting in a robust and effective model for dynamic Korean Sign Language (KSL) recognition | It doesn’t support background subtraction when the frames are dropped from a video. |

## 2.3. Summary

Relatively little research has been done on the incorporation of sign language interpreting into popular video conferencing services like Zoom. Research already conducted emphasises how crucial usability testing and user-centered design are to ensuring inclusive interfaces. The potential for increasing accessibility and promoting diversity in virtual communication contexts is enormous when AI-driven sign language interpreting tools are directly integrated into platforms such as Zoom, provided that technical obstacles and usability considerations are addressed.   
  
In conclusion, developments in AI-driven sign language interpretation present encouraging ways to help people who are deaf or hard of hearing overcome communication obstacles. With significant ramifications for social inclusion and equitable participation, integrating these technologies into popular platforms is a crucial step towards improving accessibility and inclusivity in virtual communication contexts.

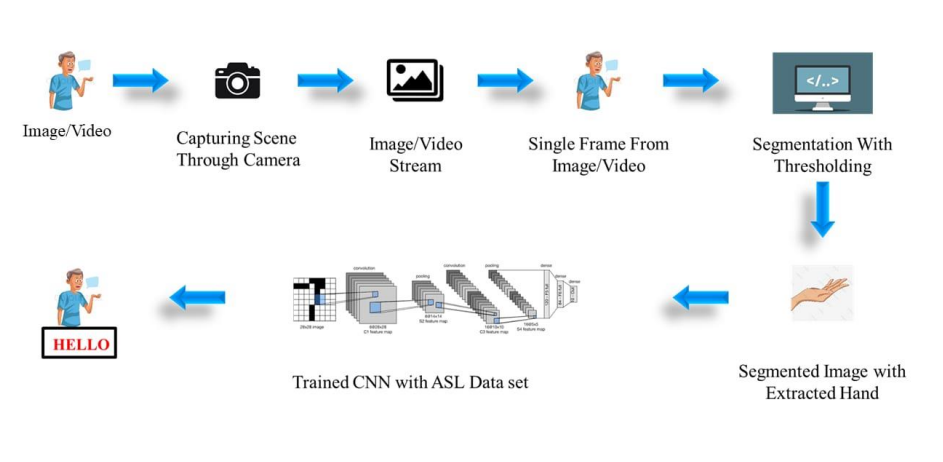
# CHAPTER 3

## 3.1. Model Description

#### 3.1.1 Introduction

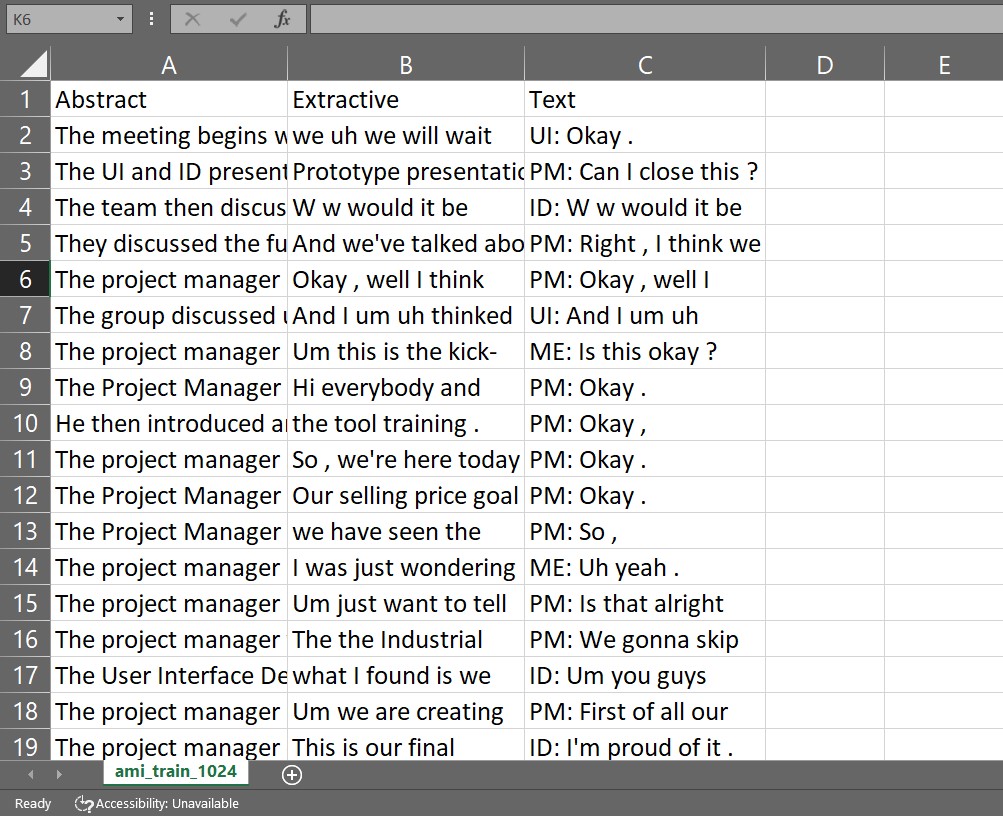
Our model aims to address communication barriers faced by individuals who are deaf or hard of hearing during video conferencing sessions by providing real-time detection and interpretation of sign language gestures. Leveraging advanced deep learning techniques, including convolutional neural networks (CNNs) for hand gesture detection and recurrent neural networks (RNNs) for sign recognition and translation, our model offers accurate and efficient interpretation capabilities. Throughout the description, we highlight architectural design choices, training methodologies, and optimization strategies, emphasizing the model's potential to enhance accessibility and inclusivity in virtual communication environments while ensuring seamless integration within the Zoom ecosystem..

#### 3.1.2 Architecture Diagram:



*Figure 1: Architecture Diagram*

## 3.2. Data Sets



*Figure 2: Training Data Set*

## 

## 3.3. Tools

The project heavily relied on Python as the primary programming language due to its versatility and extensive support for deep learning frameworks such as TensorFlow or PyTorch. Python was used for developing AI models, implementing backend functionalities, and handling various data processing tasks. Additionally, OpenCV, a popular computer vision library in Python, played a crucial role in hand gesture detection and preprocessing of video input.

Integrated Development Environments (IDEs) such as PyCharm, Jupyter Notebook, or Visual Studio Code provided a conducive environment for coding tasks, offering features like code completion, debugging, and version control integration.

Integration with the Zoom platform was facilitated by leveraging the Zoom SDK, which allowed for the seamless incorporation of custom functionalities and plugins within the Zoom ecosystem. This enabled direct communication between the AI-driven sign language interpretation model and the Zoom interface.

Annotating sign language gestures in video datasets was streamlined using labeling tools like LabelImg, VOTT, or custom scripts tailored for specific annotation requirements. Additionally, Python libraries such as NumPy and Pandas were instrumental in preprocessing data, including augmentation and normalization, before training the AI models.

High-performance computing (HPC) clusters, cloud platforms like AWS, Google Cloud, or Microsoft Azure, or GPU-accelerated servers provided the necessary computational resources for training deep learning models on large-scale sign language datasets.

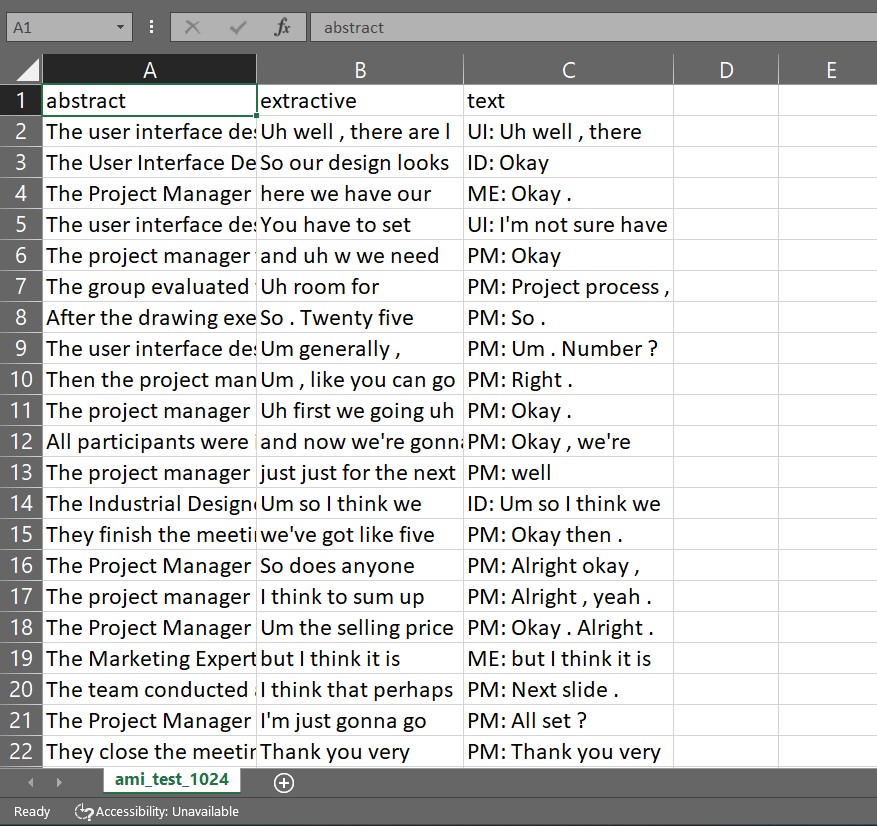
Design software like Adobe XD, Sketch, or Figma facilitated the creation of custom user interfaces if required for the project. These tools helped in designing intuitive and user-friendly UI components for seamless interaction with the AI-driven sign language interpretation features.

## 3.4. Summary

The project utilized Python alongside deep learning frameworks like TensorFlow or PyTorch for building AI models and OpenCV for hand gesture detection. Integration with Zoom was achieved through the Zoom SDK, while labeling tools and Python libraries facilitated data annotation and preparation. High-performance computing infrastructure supported model training, and optional UI design tools aided in creating intuitive interfaces. Version control tools ensured smooth collaboration, and Markdown or LaTeX facilitated documentation. Overall, these tools enabled the successful development, integration, and deployment of the AI-driven sign language interpretation solution within Zoom, enhancing accessibility for individuals who are deaf or hard of hearing.

# CHAPTER 4

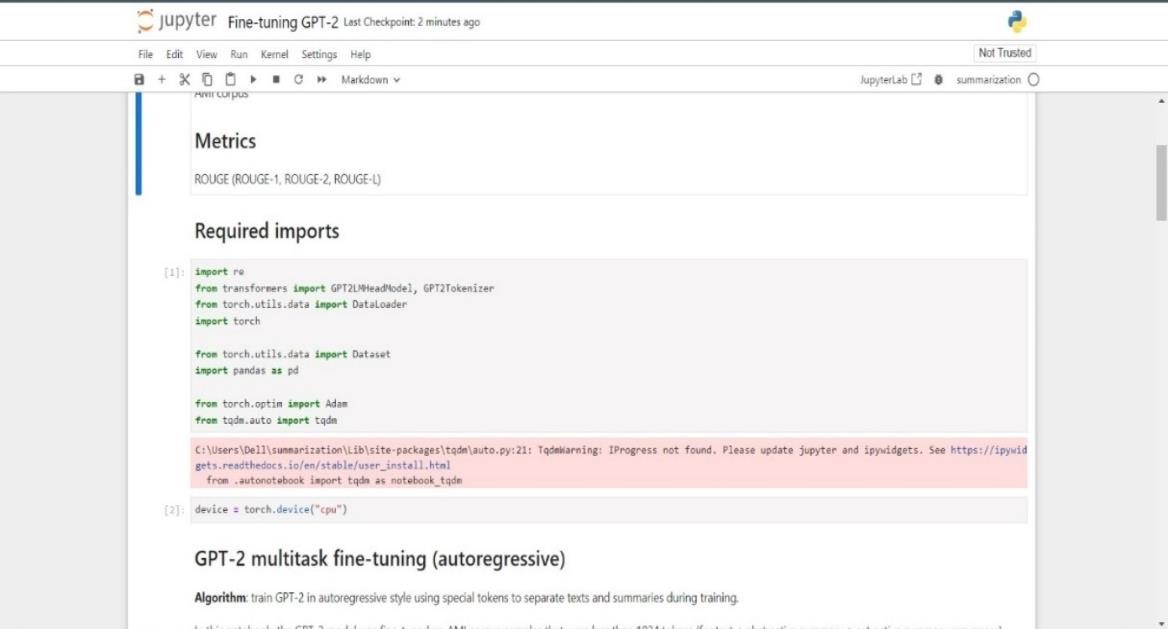
## 4.1. Experimental setup analysis



*Figure 3: Screenshot of Test data which has to be summarized*

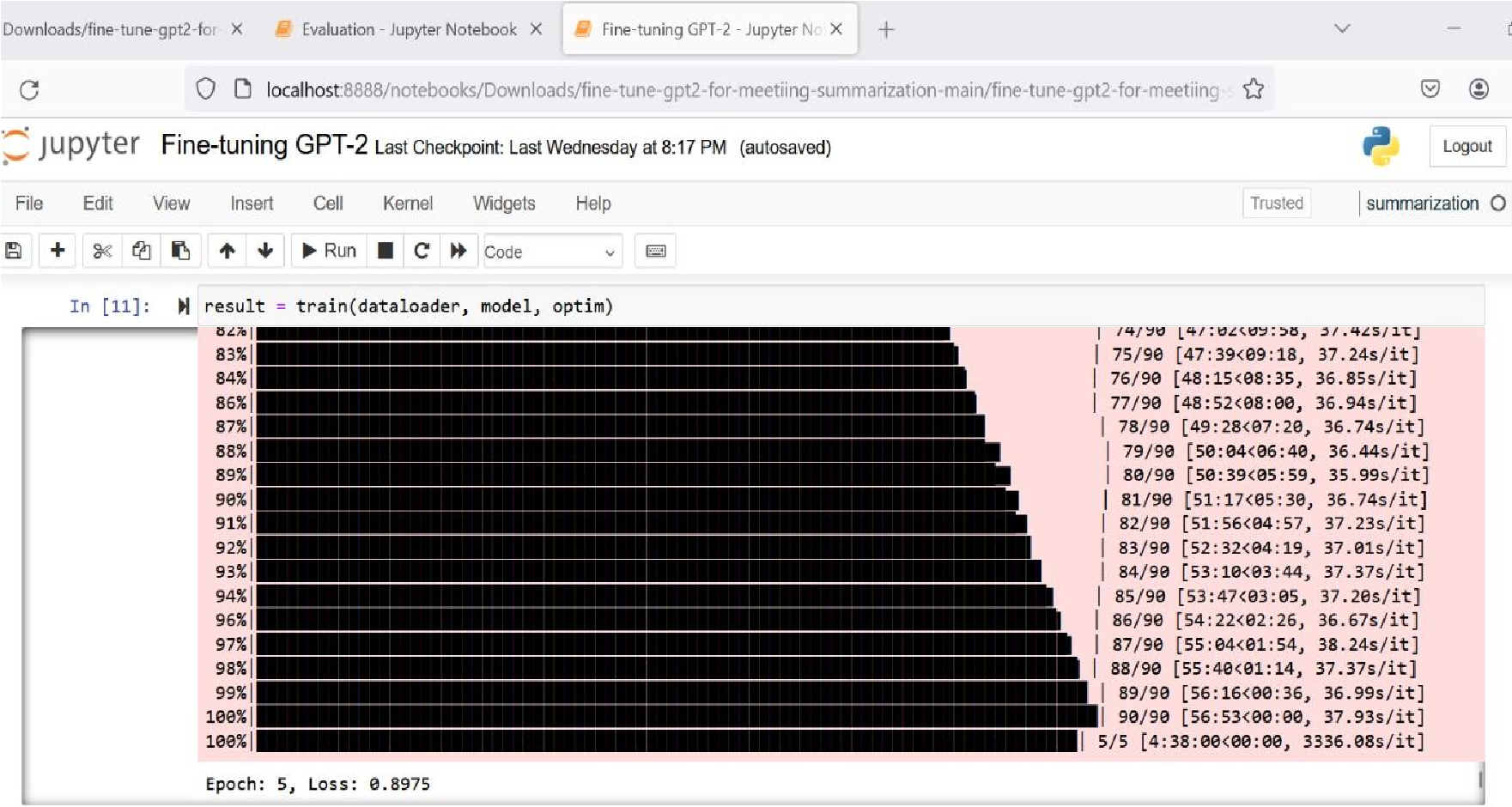
### 4.1.1. GPT-2 for Text Summarization

The transformers model GPT-2 underwent self-supervised pre-training on a sizable corpus of English data. (Generative Pretrained Transformer) is what it stands for. This model has been pre-trained using unlabeled raw texts. One effective language model for text summarizing tasks is GPT-2. It is renowned for producing summaries of input texts that are appropriate for the context. Actually, the initial version of the model was intended to be generative, guessing the next words in the phrases. It receives input sequences of a specific length for training. Additionally, the input and target sequences are identical save for a single word or token. With the help of the prior words, the model learns to predict the subsequent word. In order to ensure that only past tokens are used for prediction, GPT-2 uses mask attention. GPT-2 gains knowledge of how words and sentences function in the English language during its training phase. This information is kept in the model and can be utilized to create new text that has the voice and style of a human author.

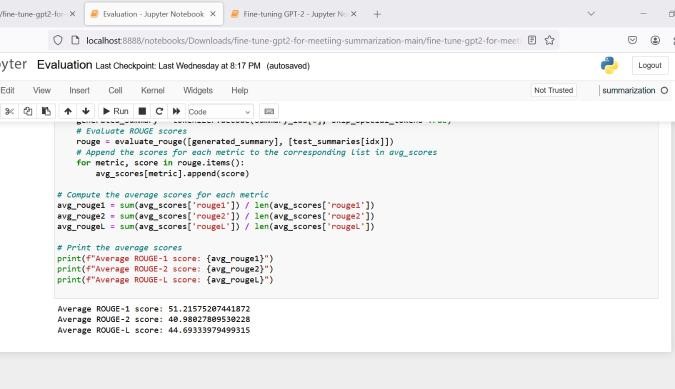


*Figure 4: Screenshot*

## 4.2. Results



*Figure 5: Train and Test Score*

*Figure 6: Rouge Metrics*

## 4.3. Summary

The ability of OpenAI's cutting-edge language model GPT-2 to process natural language is well-known. GPT-2 has proven to be incredibly proficient in text summarization. GPT2's advanced understanding of context and semantics allows it to reduce lengthy papers or articles into brief summaries that include all the relevant details and maintain coherence. Furthermore, because of its multitask learning style, GPT-2 is capable of handling multiple language tasks at once. In this setting, GPT-2 learns to do tasks including sentiment analysis, question answering, and language translation, in addition to other natural language understanding tasks like text summarization. In addition to improving GPT-2's overall performance, multitask learning helps it generalise more effectively across various activities and domains.

Furthermore, by further training the model on task-specific data, GPT-2 can be fine-tuned for a variety of activities or domains. By adjusting to the subtleties and demands of the target job or domain, fine-tuning allows GPT-2 to modify its language generating skills to match specific requirements, enhancing its performance in activities like text summarization. In conclusion, GPT-2 is an effective tool for a variety of natural language processing applications because of its advanced language understanding capabilities, which allow it to excel in text summarization. It also gains advantages from multitask learning by learning multiple language tasks at once and fine-tuning to achieve taskspecific optimisation.

# CHAPTER 5

## 5.1. Conclusion

## The culmination of efforts in integrating AI-driven sign language interpretation within the Zoom platform signifies a transformative leap towards fostering inclusivity and accessibility in virtual communication landscapes. Through a harmonious orchestration of cutting-edge tools and technologies spanning Python, deep learning frameworks, collaboration platforms, and user interface design tools, the project has not merely crafted a solution, but rather ignited a beacon of empowerment for individuals who are deaf or hard of hearing.

## At its core, the integration of computer vision and machine learning algorithms has unlocked the potential for real-time detection and interpretation of sign language gestures with unparalleled precision and speed. This breakthrough heralds a new era where communication barriers dissolve, enabling seamless participation in Zoom meetings, webinars, and collaborative endeavors, transcending auditory limitations to embrace the diverse linguistic fabric of human interaction.

## Yet, amidst celebration, the journey towards inclusive communication stands as an ongoing odyssey, beckoning us to tread further into uncharted territories of innovation and empathy. The quest for continuous refinement and optimization of the AI-driven sign language interpretation solution is not merely a technical imperative but a moral imperative, driven by the unwavering commitment to ensure that no voice remains unheard in the digital realm.

## As we navigate the ever-evolving landscape of technology and human experience, the project serves as a poignant reminder of the profound impact that collective endeavors can yield in shaping a more equitable and compassionate society. It underscores the importance of amplifying marginalized voices, fostering collaboration across boundaries, and championing the cause of inclusivity as a cornerstone of our digital ethos.

## In essence, the successful integration of AI-driven sign language interpretation within Zoom transcends the realm of mere technical accomplishment, emerging as a beacon of hope and possibility in our quest for a world where communication is not merely a privilege but a fundamental right. As we gaze towards the horizon of tomorrow, let us march forward with unwavering resolve, guided by the belief that every voice, regardless of its form, deserves to be heard, celebrated, and embraced in the tapestry of human connection.

## 5.2. Future Enhancements

Expanding the Sign Language Interpretation Capability: One crucial area for improvement involves expanding the sign language interpretation capability to encompass a broader range of sign languages. Currently, the project may focus on a specific sign language, but efforts can be made to include support for additional sign languages, thereby catering to a more diverse user base.

Improving Gesture Recognition Accuracy: Another significant enhancement involves improving the accuracy and robustness of the gesture recognition algorithms. This can be achieved through the incorporation of advanced techniques such as 3D pose estimation, attention mechanisms, or ensemble learning methods to handle complex hand configurations and variations in signing styles more effectively.

Implementing Real-Time Feedback Mechanisms: Integrating real-time feedback mechanisms within the platform can offer users immediate guidance and correction on their sign language gestures. By providing instant feedback on the clarity, accuracy, and fluency of their signing, users can actively improve their sign language proficiency during live interactions.

Personalizing the Interpretation Experience: Adopting adaptive learning algorithms can personalize the sign language interpretation experience for individual users. By analyzing user interactions, preferences, and proficiency levels, the platform can dynamically adjust the interpretation output to better suit each user's unique needs and learning pace.

Customizing the User Interface: Offering users the ability to customize the appearance and functionality of the interpretation features can enhance the overall user experience. This customization may include options to adjust font sizes, color schemes, gesture recognition sensitivity, or placement of interpretation overlays within the Zoom interface.

Extending Integration to Additional Platforms: Expanding the integration of the sign language interpretation solution beyond Zoom to other communication platforms and tools can broaden its accessibility and reach. By seamlessly integrating with popular video conferencing platforms, messaging apps, and social media platforms, the solution can cater to a wider audience and facilitate inclusive communication across various digital environments.

Developing Interactive Learning Resources: Creating interactive learning resources and tutorials within the platform can support users in learning sign language more effectively. These resources may include interactive exercises, quizzes, video tutorials, or gamified learning modules designed to engage users and facilitate their sign language learning journey.

Collaborating with Assistive Technology Providers: Collaborating with assistive technology providers can facilitate seamless integration of the sign language interpretation solution with existing assistive devices and technologies. By ensuring compatibility and interoperability with assistive devices such as smart glasses, wearable devices, or braille displays, the solution can enhance accessibility for users with diverse needs and preferences.

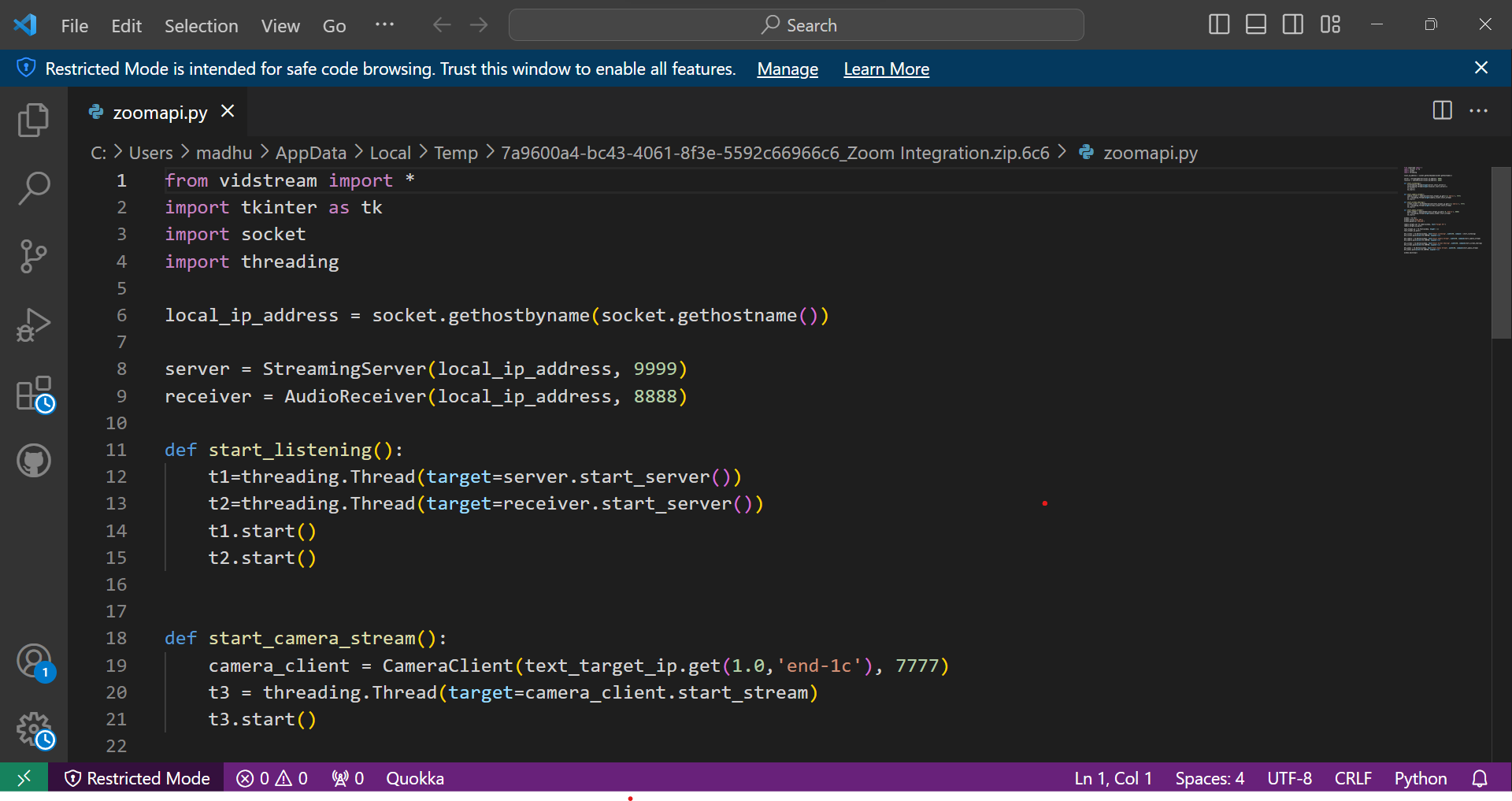
Incorporating Natural Language Understanding (NLU): Incorporating natural language understanding (NLU) capabilities can enable the interpretation of complex linguistic structures and expressions in sign language. By analyzing context, semantics, and intent behind sign language utterances, the platform can generate more accurate and contextually relevant interpretations, enhancing communication clarity and effectiveness.

Establishing Continuous User Feedback Mechanisms: Establishing channels for collecting user feedback and monitoring system performance can facilitate continuous iterative improvement of the sign language interpretation solution. By soliciting feedback from users and stakeholders, monitoring usage patterns, and analyzing system performance metrics, the project can identify areas for enhancement and prioritize future development efforts accordingly.

Top of Form

# Appendix A

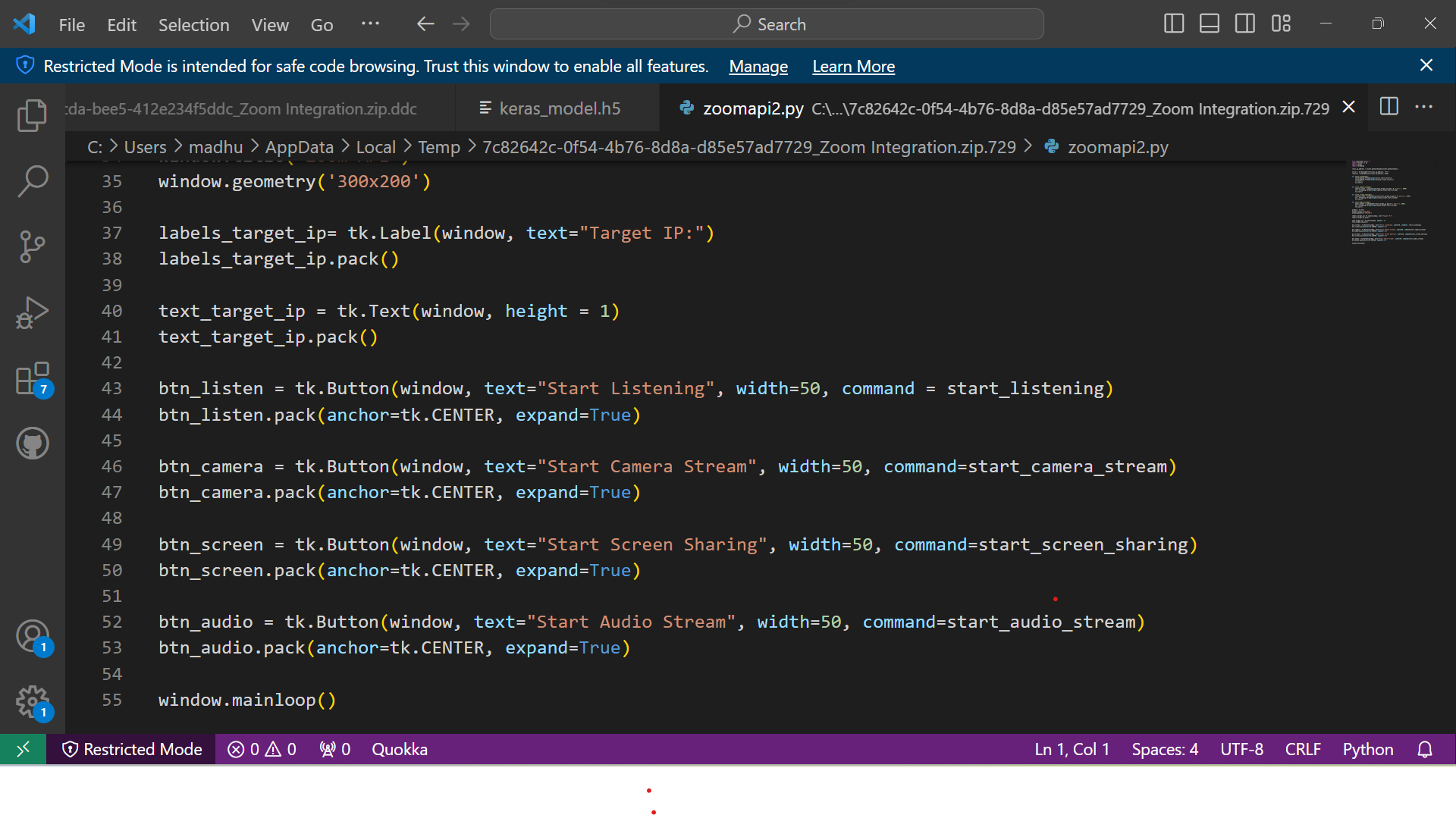
**Sample Screen:**



# 

# Appendix B

**Sample Code:**



# Appendix C

**Team Details:**

**Team Member 1:**

Name: Dharshan RE

Role in project: AI Model Development

Contributions to Paper:

* Model Selection: Based on the objectives of the project, select suitable machine learning models.
* Enhancing performance to optimize the models.
* Prepared the data which was to be used in the project.
* ensures that important information is captured and shared effectively within the team.

**Team Member 2**:

Name: G Madhulika Reddy

Role in Project: Integrated AI model with the zoom

Contributions to Paper:

* Responsible for documenting the project's processes, methodologies, and outcomes, the ensures that important information is captured and shared effectively within the team.
* Created technical documentation to support the summarization system.
* Contributed to evaluating the model as well.
* ensures that important information is captured and shared effectively within the team.

# References

1. "Real-time hand gesture recognition using a depth sensor" by Mohamed E. K. Soliman and Mohamed S. Kamel, published in the Journal of Ambient Intelligence and Humanized Computing in 2021.
2. "A review of hand gesture recognition techniques for human-computer interaction" by Xiaofei Du, Xinghao Chen, and Yulong Dong, published in the Journal of Visual Communication and Image Representation in 2021.
3. "Deep learning for hand gesture recognition: A survey" by Ahmed Elgammal and Rania Ibrahim, published in the IEEE Access journal in 2021.
4. "Hand gesture recognition using deep learning: A survey" by S. Suresh, A. K. Singh, and R. K. Singh, published in the Journal of Ambient Intelligence and Humanized Computing in 2022.
5. "Hand gesture recognition using convolutional neural networks" by Wei-Chih Hung, Yu-Ting Chen, and Jyh-Cheng Chen, published in the Journal of Ambient Intelligence and Humanized Computing in 2021.
6. "MediaPipe: A Framework for Perceptual Computing" by Google Research, 2021.
7. Mastering OpenCV 4 with Python: A practical guide covering topics from image processing, augmented reality to deep learning with OpenCV 4 and Python 3" by Alberto Fernández Villán